**EXAMEN DATA ADVANCED 2 TIN: JUNI 2020  
LITERATUURSTUDIE**

**NAMEN:** Dries Claesen & Hursit Tarcan   
**KLAS:** 2TINM  
**JAAR:** 2020  
**LECTOR:** Heidi Tans

**INLEIDING**

In dit document worden twee artikels besproken, één waarvan de datarepresentatie voldoet aan de verwachtingen en één waarbij er duidelijk te zien is dat er fouten zitten in de representatie. Eerst wordt het juiste artikel uitgebreid uitgelegd waarna we ook een samenvatting geven en uitleggen waarom we deze data representatie goed vinden. Hierna gaat de verkeerde data representatie uitgelegd worden met daarna een uitleg waarom we deze verkeerd vinden.

**DATA REPRESENTATIE – GOED  
INLEIDING**

In dit deel van het document gaat er een uitleg gegeven worden over de website met de data representatie die wij goed vinden. Omdat het de bedoeling is om een recent artikel te nemen en omdat we in deze tijd bijna alleen maar over het corona virus horen, hebben wij gekozen voor hetzelfde te doen. Wij behandelen een website waar we de corona gevallen kunnen bekijken van 215 landen.

De naam van deze website is ‘WorldOMeter’, de link naar deze website staat onderaan toegevoegd\*. Wij hebben de resultaten gebruikt die ter beschikking waren op 23 mei 2020. De hele website staat ook toegevoegd onder ‘GEBRUIKTE BRONNEN’ in dit document.

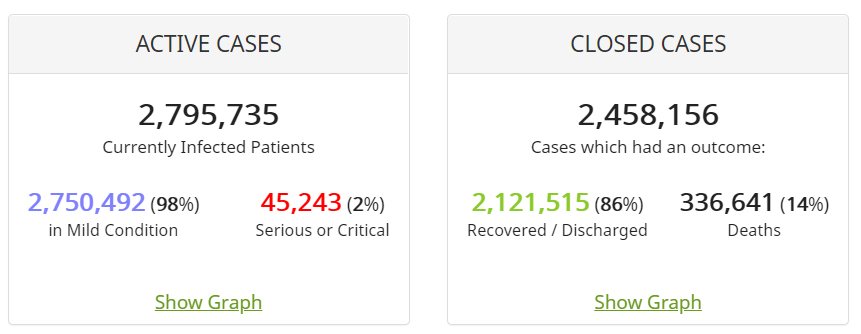
\* <https://www.worldometers.info/coronavirus/countries-where-coronavirus-has-spread/>

**DATA REPRESENTATIE – GOED  
UITLEG**

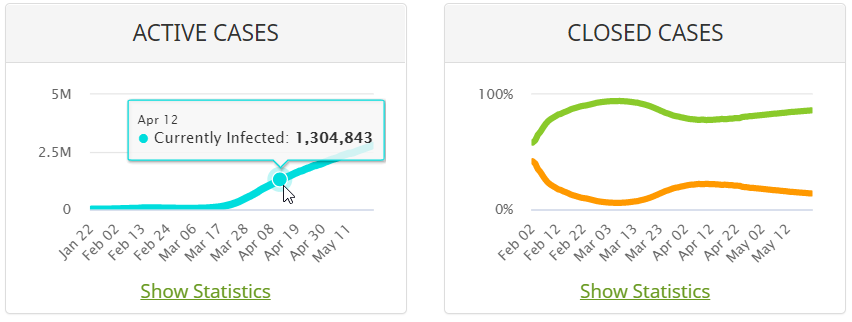
Het eerste dat we zien wanneer we op de website zitten is een overzicht van het totaal aantal corona gevallen, het totaal aantal overlijdens en het aantal mensen die hersteld zijn (zie onderstaande foto), in dit geval zijn de resultaten tot en met 22 mei 2020: 17:05. We zien het aantal gevallen op 5.253.891, het aantal overlijdens op 336.641 en het aantal dat hersteld is op 2.121.515.



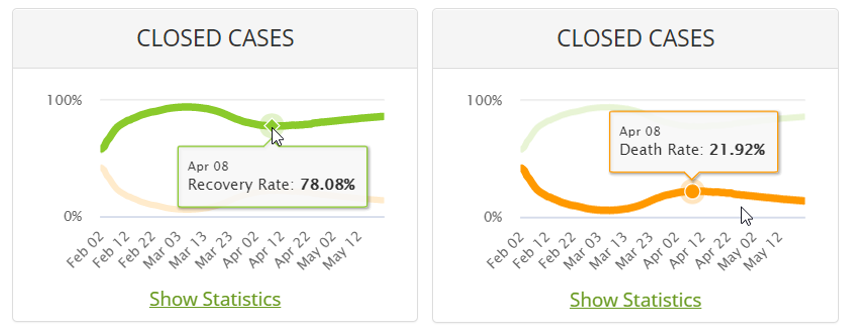
Het volgende deel van de website geeft ons een beter zicht over de huidige situatie door de actieve en de afgehandelde gevallen op te splitsen. Bij de twee delen zien we een extra opsplitsing, bij actieve gevallen is dit een onderscheid tussen de milde en de kritieke gevallen. Bij de gevallen die afgehandeld zijn is dit een splitsing tussen het aantal herstelde mensen en het aantal mensen die gestorven zijn.



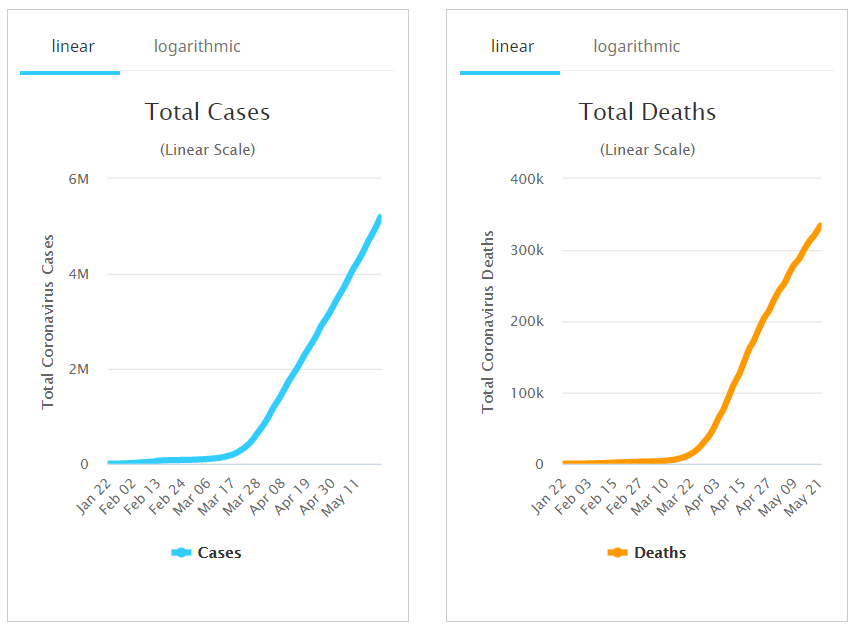
Het is ook mogelijk om deze weergave in een grafiek te tonen, we kunnen hier ook heel gemakkelijk het exacte aantal gevallen zien door gebruik te maken van onze muis. Zoals op onderstaande screenshot kunnen we dan over de grafiek gaan en het exacte nummer aflezen. In dit geval zien we het totaal aantal actieve gevallen op 12 april 2020. (zie onderstaande grafiek)



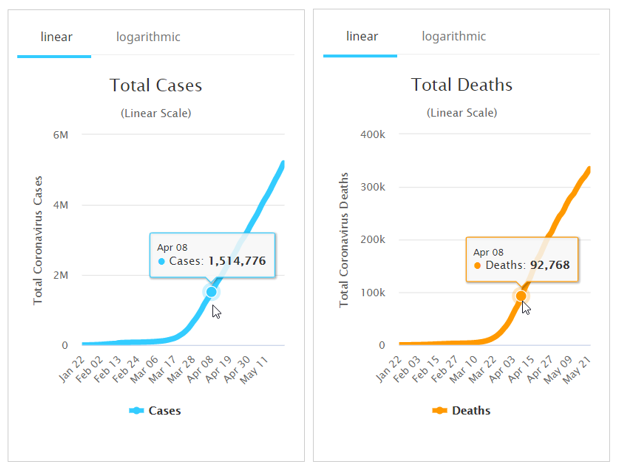
Bij de gesloten gevallen zien we ook weer dat we het exacte aantal kunnen bekijken door onze cursor te gebruiken. Op onderstaande foto’s zien we een groene en een oranje lijn. De groene lijn handelt over het percentage van het totaal aantal mensen dat hersteld is, de oranje lijn gaat over het percentage van het totaal aantal doden. Het moment dat gekozen is, is 8 april 2020. We zien hier ook heel duidelijk dat de percentages kloppen, 78.08% en 21.92% maakt samen 100%.



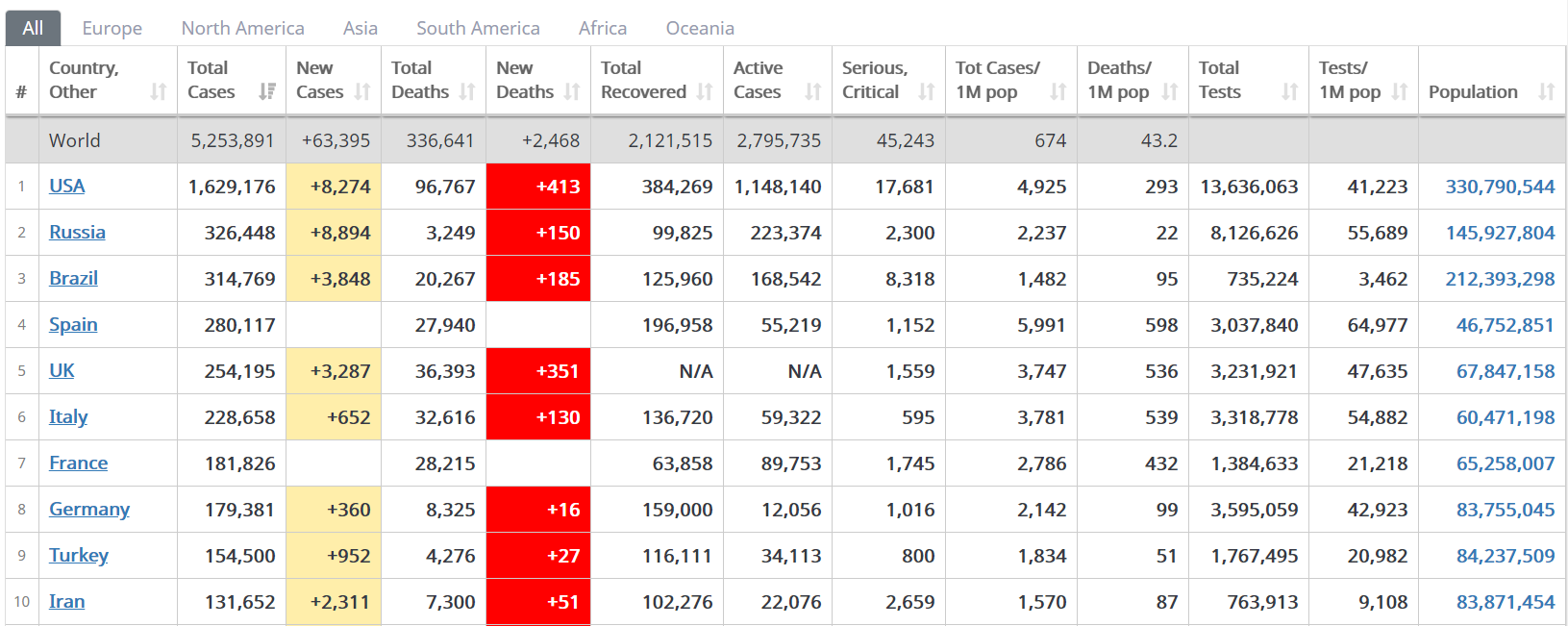
Het volgende dat we zien op deze website zijn twee andere grafieken, deze keer gaat het over het totaal aantal gevallen (blauwe lijn) en het totaal aantal doden (rode lijn). Beide grafieken tonen een verloop vanaf 22 januari 2020. We zien ook duidelijk een lineair verband tussen deze twee grafieken.



Hier kunnen we ook net zoals op de vorige grafieken het exacte aantal bekijken door onze cursor te gebruiken. Zoals op onderstaande afbeelding kunnen we dan de exacte gegevens bekijken van 8 april 2020.



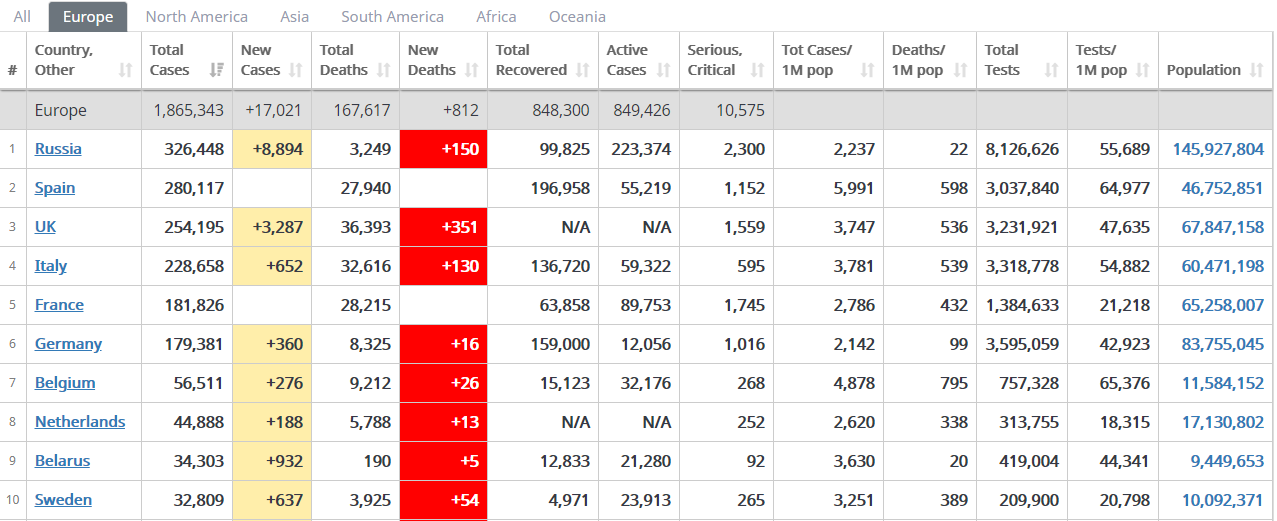
Nu komen we bij het laatste en meest interessante deel van de website. Nu zien we een tabel met 215 landen waarvan de corona gegevens zijn verzameld. De bronnen van deze gegevens komen voor het grootste deel van officiële organisaties (o.a. ministeries). Omdat de lijst heel groot is hebben we ervoor gekozen om hier enkel een screenshot te tonen van de top 10 landen waar deze pandemie voor de meeste gevallen heeft gezorgd. De volledige lijst is onderaan dit document te vinden.

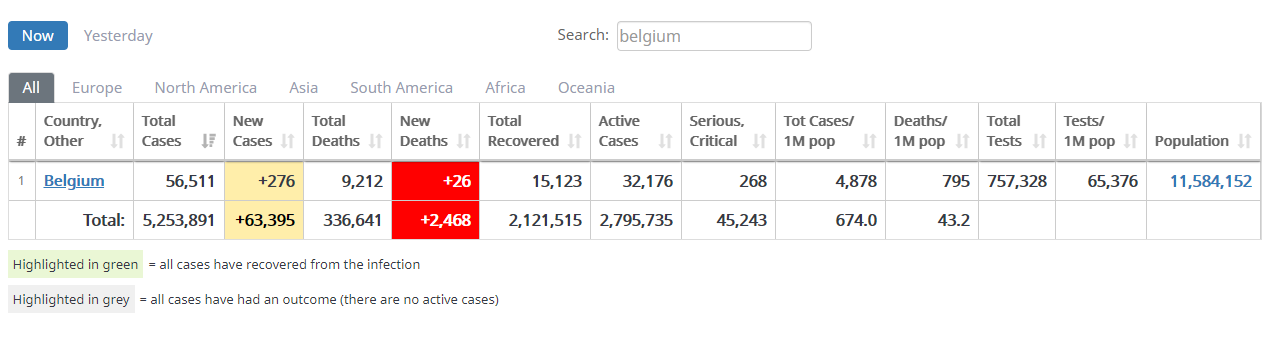


Als we deze tabel opsplitsen en helemaal analysen dan zien we dat het eerste deel de naam van het land in kwestie bevat. Als voorbeeld nemen we het land dat op de eerste plaats staat, namelijk USA. Als we een plaats naar echts opschuiven dan zien we dat er in totaal 1.629.176 corona gevallen zijn. Daarlangs zien we dan ook het aantal nieuwe gevallen (deze zijn weergeven in een andere kleur), in dit geval zijn er 8.274 nieuwe gevallen. Bij het aantal doden zien we ook een opsplitsing tussen het totaal aantal en het nieuwe aantal. Hier zien we een totaal aantal van 96.767 en in een rode kleur zien we dan dat er 413 doden zijn bijgekomen.

Het volgende deel gaat dan over het aantal mensen die hersteld zijn, dit zijn er 384.269. We zien daarna ook het aantal actieve gevallen met één extra opsplitsing, namelijk het aantal kritieke gevallen. Verder krijgen we ook andere interessante informatie te zien zoals het aantal gevallen en doden per 1 miljoen mensen, we zien ook het totaal aantal testen dat is gedaan en het aantal dat is gedaan per 1 miljoen mensen. Als laatste zien we dan ook het cijfer van de totale populatie van het land in kwestie, dit is ook heel interessant om te zien.

Verder is het ook mogelijk om de tabel op verschillende manieren te filteren. Bijvoorbeeld alfabetisch op land, landen met meeste doden eerst, landen met meeste populatie, filteren op een specifiek continent of land en veel meer. Op onderstaande afbeelding zien we dat het gaat over de top 10 landen met meeste gevallen in Europa. Op de foto daar onder zien we de cijfers van belgië.



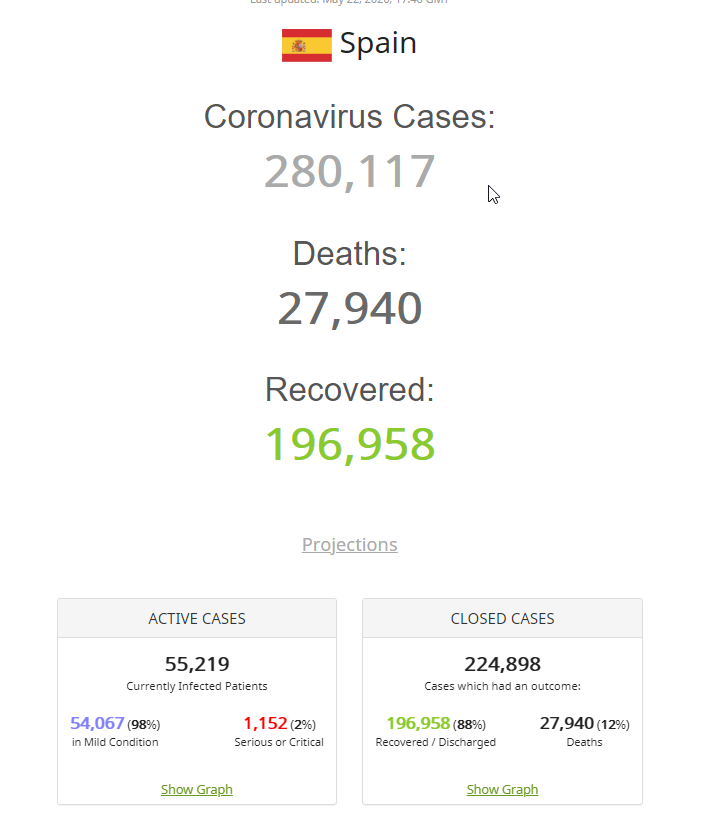


**DATA REPRESENTATIE – GOED  
SAMENVATTING**

We vinden deze website goed omdat we een goede combinatie zien tussen meerdere elementen. We kunnen totale cijfers zien maar we zien ook een juiste opsplitsing van deze totale getallen. Bijvoorbeeld bij het totaal aantal gevallen, hier zien we ook een opsplitsing in het aantal actieve en het aantal gesloten gevallen. Het is ook handig dat de grafieken langs elkaar staan, dit maakt het heel makkelijk om alles met elkaar te vergelijken.

Het is ook leuk dat we het exacte cijfer kunnen bekijken wanneer we over de grafiek gaan, dit zorgt ervoor dat we geen fouten kunnen maken bij het aflezen van de data. In de frequentietabel van de 215 landen is het handig dat we meer informatie kunnen zien dan enkel het aantal gevallen, de totale populatie en het aantal testen die gedaan zijn zijn ook heel interessant om af te lezen. De filters die we kunnen gebruiken zijn ook goed uitgewerkt, dit zorgt ervoor dat we alles van groot naar klein en andersom kunnen bekijken en vergelijken. Het werken met verschillende kleuren en interne links is ook zeer handig, door de interne links kunnen we een land uitgebreid analyseren door op de naam te klikken, in onderstaande afbeelding zien we de gegevens van Spanje.

De grafieken zijn in het algemeen ook heel duidelijk te lezen, we zien dat de x en y as geen namen hebben maar het is niet moeilijk om af te lijden waarover het gaat. Verder kunnen we ook bevestigen dat de getallen van officiële bronnen komen, onderaan op de website zien we een complete bronnenlijst waarbij we de organisaties kunnen bekijken waar deze cijfers vandaan komen. Wij willen zeker niet zeggen dat deze cijfers 100% juist zijn, maar de kans dat deze bronnen juiste gegevens doorgeven is veel hoger dan gegevens die we krijgen via o.a. Facebook of Twitter. Ook zien we dat de percentages in de grafieken juist in elkaar zitten, waarbij we nooit boven de 100% uitkomen wanneer we kijken bij het percentage van herstelden t.o. het percentage van gestorven mensen.



**DATA REPRESENTATIE – SLECHT   
INLEIDING**

In dit deel van het document gaat er een uitleg gegeven worden over de tweet met de data representatie die wij niet goed vinden. Het gaat hier over een tweet van Global Investigative Journalism Network. De tweet gaat over een cirkeldiagram die getoond werd tijdens een uitzending van de Amerikaanse TV zender WCVB. In het diagram is een fout te zien die veel mensen snel opmerken. Deze zending is gebeurd in maart 2020.

**DATA REPRESENTATIE – SLECHT   
UITLEG**

De zending gaat over de grootste zorgen die de Amerikaanse bevolking heeft door het corona virus. De top drie problemen volgens WCVB zijn:

* Het virus zelf krijgen
* Iemand van de familie die het virus krijgt
* Een economische crisis door het virus

Met deze problemen is er helemaal niks mis, maar wanneer we naar het diagram zelf kijken. Dan zien we dat er een groot probleem is. Elke persoon die ooit geleerd heeft over een cirkel diagram weet ook dat het totaal altijd 100% moet zijn, wat hier helemaal niet het geval is. Niet enkel zijn de percentages fout, maar ook de vakken die gekleurd zijn zijn helemaal niet in verhouding met het percentage dat erbij staat. Bijvoorbeeld op een normaal cirkeldiagram zou het rode vak ongeveer 28% moeten zijn, hier is het 48% (48% zou eigelijk bijna de helft moeten zijn van de hele cirkel).

Verder hebben we ook een kleine opmerking over de volgorde van de legende, dit is helemaal geen ernstige fout maar het speelt wel een rol in hoe makkelijker het diagram te bekijken is. Zoals we zien begint de legende met een rode kleur, maar als we het diagram bekijken zullen de meeste mensen eerst focussen op het bovenste deel en daarna met de klok meegaan om het diagram volledig te bekijken. Als de persoon in kwestie hier kijkt bij het bovenste (dus oranje) en daarna gaat kijken in de legende, dan staat de uitleg op plaats twee, dit is helemaal geen grote fout maar het zou veel gemakkelijker zijn als de legende op de volgende manier was geplaatst. Eerst ‘Oranje’, daarna ‘geel’ en als laatste ‘rood’.



**GEBRUIKTE BRONNEN**

